

Веселова Н. А. Заглавие литературно-художественного текста: онтология и поэтика: Дис. ... канд. филол. наук. Тверь, 1998.

Мильчин А. Э., Чельцова Л. К. Справочник издателя и автора: Редакционно-издательское оформление издания. М., 1999.

Мирошникова О. В. Итоговые книги конца XIX века и книготворчество русского символизма // Русская классика: между архаикой и модерном. СПб., 2002.

Орлицкий Ю. Б. Минимальный текст (удертон) в составе заголовочно-финального комплекса стихотворного текста // Поэтика заглавия. Москва; Тверь, 2005.

Толстых Г. А. Прижизненные стихотворные сборники русских символистов: книготворчество поэтов // Книга: Исследования и материалы. Сб. 62. М., 1991.

Тюпа В. И. Литература как род деятельности: теория художественного дискурса // Теория литературы: В 2 т. Т. 1. Теория художественного дискурса. Теоретическая поэтика. М., 2004.

*Статья поступила в редакцию 14.06.07.*

**М. Ю. Мухин**

## **ЛЕКСИЧЕСКИЕ УНИВЕРСАЛИИ ХУДОЖЕСТВЕННОГО ТЕКСТА: СТАТИСТИКА И ИДЕОГРАФИЯ\***

Рассматривается одна из возможностей приложения количественных методов к филологическому анализу художественного текста. На основании выборки из романов XX в. формируется список наиболее частотных слов (имен существительных и глаголов), обозначенных как лексические универсалии художественного текста. Выделяются тематические группы данных слов, отмечается антропологичность универсальной лексики и относительность ее репрезентативности при описании конкретного идиостиля. Намечен подход к исследованию индивидуально-авторской идеографии, основанный на строго формализованной выборке материала.

Для многих филологических исследований характерны такие формулировки, как «Это явление часто встречается в рассматриваемом тексте», «Следует отметить также и это...» и т. п. Всегда хочется спросить: насколько часто? Почему следует отметить? Но даже если выводы, касающиеся специфики художественного текста, опираются на четкую статистику, ее применение может оказаться неудачным. Приведем типичный пример без ссылки на источник: *Для изучаемого писателя характерны высокочастотные слова **идти** и **говорить**. Из этого следует, что ситуации движения и речи в текстах писателя связаны с индивидуально-авторской картиной мира и являются более важными, чем для других авторов.* Конечно, частотная выборка объективирует исследование, однако подобные идиостилевые рассуждения сомнительны: основная лексика движения и речи характерна для любого художественного текста, является универсальной. Любой текст отражает свой фрагмент действительности, но интуитивно понятно, что в основных номинациях предметов и явлений мира эти фрагменты должны пересекаться. Возникает закономерный вопрос: какие слова (группы слов) следует отсека-

\* Исследование поддержано грантом Президента РФ МК-337.2007.6.

при идиостилевом анализе, если исследование строится только на основании критерия частотности?

Допустим, что есть группы лексики, закономерно встречающиеся практически в любом тексте. Они и должны в первую очередь отражать приоритеты текстовой номинации и базовые (не индивидуальные, а общезначимые) концепты, которые автор неосознанно репрезентирует в произведении. В этом смысле интересно не просто выявить частотную лексику, а сгруппировать ее в некие семантические множества.

Подтвердить эту мысль и выстроить объективный список наиболее важных понятий, отражаемых в тексте, может только сопоставительный частотный анализ, проведенный на большом массиве словоупотреблений. При этом логично сопоставлять в одной группе тексты, написанные приблизительно в одно время.

Сходные проблемы решает в своих работах В. С. Баевский на материале русской поэзии [см.: Баевский, 2001, 192–217]. Поэтический текст имеет свою специфику: его малая протяженность дает автору возможность в большей степени отразить «доктрину» собственного, личностного восприятия мира (например, употребить или не употребить несколько раз слово *душа*). В итоге поэзия (и в этом ее эстетическая стезя) должна давать больше искажений в сравнении с более близкой к естественной речи прозе. Только в этом смысле можно говорить о том, что результаты частотного анализа прозы объективнее.

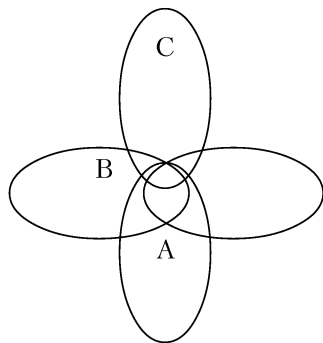
Настоящая статья отражает некоторые результаты лингвостатистического исследования, проводимого нами на материале 16 прозаических произведений М. Булгакова, В. Набокова, А. Платонова, М. Шолохова, содержащих в сумме около 1,5 млн словоупотреблений<sup>1</sup>. Для рассмотрения взяты самые известные русские романисты XX в., произведения которых отражают состояние языка и текста, характерное для первой половины прошлого века.

Материал (знаменательные слова) ограничен нарицательными именами существительными и глаголами. В будущем планируется присоединить к ним и другие знаменательные части речи. Концептуальное соотнесение частотности служебных слов и местоимений, на наш взгляд, не может дать существенных результатов (что скажет о концептуальных предпочтениях и идиостиле автора  $X$  несколько большая встречаемость союза «и» по сравнению с произведениями автора  $Y$ ?). Служебные слова и личные местоимения всегда суперчастотны, а их соотнесение плодотворно используется при решении обратной проблемы — атрибуции текстов. Создатели моделей определения авторства могут позволить себе не интересоваться семантикой.

Из каждого текста выбрано по 300 самых частотных лексем, соответствующих условию. Проведено сопоставление списков. Гипотеза заключается в следующем. Частотные ряды лексем, характеризующие произведения четырех авторов, распределяются в следующие группы (см. схему): группа *A* — слова, часто встречающиеся в произведениях всех писателей; группа *B* — слова, объединяющие 2–3 писателей; группа *C* — лексика, характерная только для одного автора.

---

<sup>1</sup> Выборка осуществлена по произведениям «Белая гвардия», «Жизнь господина де Мольера», «Театральный роман», «Мастер и Маргарита» М. Булгакова; «Король, дама, валет», «Защита Лукина», «Камера обскура», «Дар» В. Набокова; «Сокровенный человек», «Чевенгур», «Котлован», «Ювенильное море» А. Платонова; «Тихий Дон», «Поднятая целина», «Они сражались за Родину», «Судьба человека» М. Шолохова.



Проведя сопоставительный анализ, распределяющий лексический массив по количеству произведений, в которых слова оказываются частотными, выберем лексику группы А и постараемся дать идеографическую характеристику выявленным словам. При выделении семантических групп используются классификации, принятые в идеографических словарях, изданных под ред. проф. Л. Г. Бабенко лексикографической группой «Русский глагол» [Толковый словарь, 1999; Большой толковый словарь, 2005].

Исследование показывает, что лексические универсалии текста действительно существуют. Рассмотрим наиболее употребительные существительные и глаголы.

### Универсальные имена существительные

Полученный список (41 существительное) можно тематически систематизировать следующим образом:

1) наиболее общие обозначения человека: *человек (люди), женщина*. Слово *мужчина* почти никогда не является частотным. Заметим, что в русском языке, как и во многих других, слово *человек* в разных контекстах является синонимом к слову *мужчина*. Ср.: *Вас какой-то человек спрашивает. Человек остановился у двери* — едва ли можно решить, что речь идет о женщине. Слово *мужчина* характеризуется более узкой семантической и стилистической сочетаемостью;

2) термины родства: *мать, жена*; прочие (*отец, сестра, брат, муж* и т. п.) существенно отстают;

3) органы тела, внешние признаки человека: *голова, рука, лицо, глаз, нога, сердце*; можно объединить с ними и слово *голос*;

4) время и его промежутки: *время, год, день, ночь, вечер, утро, час, минута*;

5) пространство и его координаты: *мир, земля, место, сторона, город, улица, двор, дом*; сюда же примыкают слова, обозначающие части жилища: *комната, окно, дверь, пол*, а также мебель: *стол*;

6) слово *жизнь* и наименования наиболее важных обобщенных результатов деятельности человека: *мысль, слово, дело*;

7) особняком стоят слова *вода, свет, Бог, солнце, конец*. В силу очевидной частотности разных значений этих слов (проблема многозначности здесь не решается) они не могут быть причислены, например, к группам «пространство», «жизнь» и т. д.

Итак, слова тематических групп «человек», «время» и «пространство» оказались самыми важными: они задают основные координаты не только текста, но и человеческой жизни вообще. Любопытно, что они с такой последовательностью проявляются в каждом художественном тексте. Речь идет о показателях от 0,04 до 0,64 % всех слово-

форм, что в крупном произведении составляет не одну сотню словоупотреблений, например, *человек (люди)* («Чевенгур») — 741 случай, *рука* («Мастер и Маргарита») — 356, *глаз* («Белая гвардия») — 238, *год* («Дар») — 229.

### Универсальные глаголы

Группу глаголов, являющихся частотными во всех текстах, составили 53 слова. Члены видовых пар подсчитывались вместе, т. е. условно принимались за одну единицу. Специфика глагольной многозначности не дает возможности без определенных погрешностей провести идеографическую стратификацию, тем не менее ряд семантических направлений можно указать с достаточной степенью точности.

Универсальными являются следующие слова:

1) глаголы, которые могут употребляться в функции отвлеченной и полузнаменательной связок: *быть, становиться/стать*;

2) модальные глаголы: *мочь, хотеть*.

Далее — более конкретные, обозначающие значимые процессы глаголы:

3) движения: *идти, пойти, ходить, приходить/прийти, уходить/уйти, проходить/пройти, подходить/подойти, входить/войти, выходить/выйти, падать/упасть*;

4) перемещения: *поднимать/поднять, снимать/снять*;

5) речи: *отвечать/ответить, спрашивать/спросить, говорить/сказать, писать*;

6) интеллекта: *знать, узнавать/узнать, понимать/понять, думать, подумать*;

7) зрительного восприятия: *видеть, увидеть, глядеть/глянуть, смотреть, посмотреть*;

8) местонахождения, положения в пространстве: *стоять, лежать, сидеть, садиться/сесть, оставаться/остаться*;

9) получения в свое распоряжение, владения, передачи объекта: *иметь, получать/получить, брать/взять, давать/дать, держать*.

Вне перечисленных групп следующие многозначные слова: *делать/сделать, жить, улыбаться/улыбнуться, начинать/начать, находить/найти, ждать, приходиться/прийтись, попадать/попасть, помогать/помочь, оставлять/оставить, показывать/показать, успевать/успеть, открывать/открыть*.

Частотность универсальных глаголов дает текстовое подтверждение идее Л. Г. Бабенко о том, что «в большей степени способны номинировать концепты процессуально-событийного мира базовые глаголы-идентификаторы с социально-обобщенным типом значения» [Бабенко, 1999, 35]. Несмотря на то что сферы движения, интеллекта, речи и т. п. по-разному выражены в произведениях четырех писателей, и в этом проявляются идиостилевые черты, есть, безусловно, универсальные процессы типа *отвечать, видеть, давать* и т. п. Частотные глаголы так же, как и существительные, антропологичны по семантике. Обозначаемые процессы связаны с восприятием мира человеком, нахождением в пространстве, общением, социальными отношениями.

В чем практический смысл выявления этого списка? Во-первых, частотные словари пока не дают развернутой статистики относительной встречаемости слов в конкретных художественных произведениях. Многие корпуса русских текстов находятся сегодня в стадии становления и содержат большой неразмеченный сектор<sup>2</sup>. Часто словари

<sup>2</sup> Таков, например, активно развивающийся проект «Национальный корпус русского языка» ([www.ruscorpora.ru](http://www.ruscorpora.ru)) и ряд других.

строятся на нерасчленном массиве, содержащем тексты разных функциональных стилей, в особенности публицистику. Рассмотренные слова — самые яркие примеры лексических совпадений. При этом огромные пласты частотных слов объединяют не десять, а двух-трех авторов. Их тоже нельзя называть индивидуально-авторскими.

Не очевидно ли, что в каждом тексте отражается человек, его представления о времени и пространстве и основные физиологические, социальные, природные процессы? Следовало ли проводить такой анализ, если его результаты с поверхностной точки зрения интуитивно понятны? Думается, что следовало. Интуитивное читательское и лингвистическое представление, с которым, в общем, не расходятся полученные данные, не может дать четкой статистической картины; более того, оно часто подводит филолога, влияя на его гуманитарный взгляд и делая интерпретацию текста субъективной. Кроме того, многие близкие по значению слова, называющие одни и те же объективные реалии, количественно по-разному представлены в текстах. Одни из них попадают в «пик частотности» (глагол *говорить*), другие — нет (например, глагол *вскрикивать* у Булгакова, *пояснять* у Набокова, *гутарить* у Шолохова), отражая таким образом авторскую языковую картину мира.

Итак, можно говорить о том, что универсальную лексику (как бы частотна в отдельном романе она ни была) не стоит привлекать без контекстологического анализа для характеристики отдельного идиостиля. Например, только частотность существительного *город* в романе «Белая гвардия» или глагола *ходить* в романе «Тихий Дон» никак не может свидетельствовать о концептуальной значимости соответствующих понятий для названных произведений. На этот вопрос может ответить исключительно филологическая интерпретация самих контекстов. Нельзя, вероятно, привлекать подобные слова и для решения проблемы атрибуции текста. Если же семантическому анализу предшествует частотная выборка, она должна быть обязательно сопоставительной. Такая выборка выводит исследователя текста за рамки одного произведения или творчества одного автора и позволяет отсеять индивидуальные явления от общих. Современные корпуса текстов и частотные словари дают филологам новые возможности в этом плане. В идеале частотное, контекстологическое и концептуально-текстовое направления изучения текста должны не исключать, а гармонично дополнять друг друга.

Дальнейшая логика исследования приводит нас к анализу индивидуальной лексики, характерной для каждого конкретного автора, однако это материал для отдельной работы. Интересно было бы сопоставить полученные данные с частотностью лексики, которую демонстрирует русская классика XIX в., а также современная литература.

---

Бабенко Л. Г. Базовые концепты процессуально-событийного мира (постановка проблемы) // Русская глагольная лексика: Денотативное пространство. Екатеринбург, 1999.

Бабевский В. С. Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. М., 2001.

Большой толковый словарь русских существительных: Идеогр. описание. Синонимы. Антонимы. М., 2005.

Толковый словарь русских глаголов: Идеогр. описание. Английские эквиваленты. Синонимы. Антонимы. М., 1999.

Статья поступила в редакцию 20.04.07.